DOCUMENT RESUME

ED 480 210 UD 035 840

AUTHOR	David, Jane L.; Shields, Patrick M.
TITLE	When Theory Hits Reality: Standards-Based Reform in Urban Districts. Final Narrative Report.
INSTITUTION	SRI International, Arlington, VA.
SPONS AGENCY	Pew Charitable Trusts, Philadelphia, PA.
PUB DATE	2001-08-00
NOTE	58p.; Written with the assistance of Daniel C. Humphrey and Viki M. Young.
AVAILABLE FROM	For full text: http://www.sri.com/policy/cep/edreform/pew.html#reports.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Standards; *Accountability; Change Strategies; Curriculum Development; Educational Assessment; *Educational Change; *Educational Improvement; Elementary Secondary Education; Student Evaluation; Teaching Methods; *Urban Schools
IDENTIFIERS	Low Performing Schools; Reform Efforts

ABSTRACT

In 1996, the Pew Charitable Trusts gave 4-year grants to seven urban school districts to help implement standards-based systemic reform. Several times each year, researchers conducted interviews to track progress in their reform efforts. This report synthesizes findings across all 4 years of the Pew Network. Overall, Network sites worked hard to put standards and assessments in place and attempted to create sustainable accountability systems. They all encountered difficulties, including considerable turnover of top leadership and changes in the direction of state policies. A central finding was that the core components of standards-based reform (standards, assessments, and accountability) did not do a very good job of communicating high expectations of students, providing information to guide instructional improvement, or motivating widespread instructional change beyond test preparation. Where there were districtwide changes in instruction, district leaders had communicated a clear set of expectations for instruction through curriculum adoptions or other frameworks, and they had backed up their expectations with intensive professional development. Districts that significantly expanded and revamped their professional development had to make major shifts in priorities, budgets, and staffing. Districts and states dedicated disproportionately more resources and attention to identifying failures than prevention or assistance. (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 210

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Urahn
Pew Charitable Trusts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

August 2001

WHEN THEORY HITS REALITY: STANDARDS-BASED REFORM IN URBAN DISTRICTS

Final Narrative Report

Prepared by:

Jane L. David, Bay Area Research Group
Patrick M. Shields, SRI International

With the assistance of:

Daniel C. Humphrey, SRI International
Viki M. Young, SRI International

SRI Project P07270
Reference Log No. 96000418-000

Prepared for:

Susan Urahn
The Pew Charitable Trusts
One Commerce Square
2005 Market Street, Suite 1700
Philadelphia, PA 19103-7017

BEST COPY AVAILABLE

UD 035 840

EXECUTIVE SUMMARY

In 1996, The Pew Charitable Trusts gave four-year grants to seven urban school districts to assist their efforts in implementing standards-based systemic reform. Called the Pew Network for Standards-Based Reform, the seven districts are: Christina, Delaware; Community District 2, New York City; Fayette County, Kentucky; Pittsburgh, Pennsylvania; Portland, Oregon; San Diego, California; and Yonkers, New York. The Trusts' primary goal in funding these districts was to test the theory and assumptions behind a standards-based systemic approach to reform. Several times each year, an evaluation team conducted interviews in the districts to track progress in their reform efforts in order to answer a set of broader questions: What does the theory look like in practice? Do the presumed relationships hold up? Do they result in higher achievement for all students? This fourth and final evaluation report synthesizes findings across all four years of the Pew Network.

The theory of standards-based reform takes the creation of three components—ambitious standards, aligned assessments, and accountability—as the starting place for increasing student achievement. The reasoning is that these reform components will communicate clear and high expectations for students, guide changes in practice, and motivate educators to improve. In exchange for accountability, schools are to have the flexibility to determine for themselves how best to increase student performance. The theory holds that these reform components, coupled with professional development in support of teaching to standards and more learning time for students who need it, will motivate and prepare teachers to implement instructional practices that will enable all students to reach high standards.

Although this logic may seem straightforward, a shift from norms to high standards for all students presumes a very different set of beliefs and roles than currently exists in most districts. Promoting teaching practices designed to help *all* students reach ambitious standards runs counter to widely shared beliefs about the nature of learning and about the abilities of many students, especially poor and minority students. In this light, creating a standards-based system presents a daunting, transformational challenge for districts.

Key Findings

We found that, over the four years of the grant, the Pew Network sites have worked hard to put standards and assessments in place and have attempted to create sustainable accountability systems. At the same time, all the sites have encountered strong headwinds, including

considerable turnover of top leadership and changes in the direction of state policies. States have become major players in standards, assessments, and accountability, but their choices are not necessarily consistent with district directions. As a consequence, the Network districts have had to contend with multiple sets of standards—some more ambitious than others—and changing assessment and accountability systems, most of which rely on norm-referenced tests.

A central finding of our evaluation is that the core components of standards-based reform—standards, assessments, and accountability—do not play their intended roles well. They do not do a very good job of communicating high expectations for students, providing information to guide instructional improvement, or motivating widespread instructional change beyond test preparation.

High-stakes accountability does motivate educators to avoid sanctions. However, the goal becomes one of raising test scores, which typically results in less ambitious teaching, especially for low-performing students. Only when the assessments encourage more ambitious teaching—for example, by asking for written arguments or applications of knowledge—did we find teachers attempting changes beyond practicing test-like multiple-choice items.

Even when teachers are motivated to adopt more effective practices, standards and assessments are not sufficient to help them do a better job. Increasing the amount of writing activities in the classroom, for example, is not the same as teaching writing better. Improvements in teaching practice do result when teachers have a clear idea of what effective instruction looks like, together with sufficient professional development and support to learn new ideas and put new practices in place.

Where we documented districtwide changes in instruction, district leaders had communicated a clear set of expectations for instruction through curriculum adoptions or other frameworks. And they had backed up their expectations with intensive professional development on teaching specific content (reading or mathematics) and ongoing school-based assistance. In fact, the only reform effort across the districts that clearly resulted in student achievement gains had clear instructional expectations, supported by extensive professional development, over a period of several years (predating the Pew Network).

Districts that significantly expanded and revamped their professional development had to make major shifts in priorities, budgets, and staffing. Because these changes have tremendous financial and political ramifications, however, they are not attractive to many district leaders. Moreover, when districts make such decisions centrally and dictate the instructional agenda for

all schools, they run up against traditions of school autonomy and the theoretical notion that schools gain flexibility in exchange for accountability.

We also found that districts, and their states, dedicate disproportionately more resources and attention to identifying failures than to prevention or assistance. Few of the districts, or their states, have made substantial investments in preventing failures from occurring through early intervention. Nor have many invested in intensive strategies to assist failing schools and students.

Conclusions

These findings demonstrate that the ideas embodied in standards-based reform are exceedingly difficult to realize in urban districts. They do not negate the overall idea of standards—in fact, the exceptions in our sites can be interpreted as support for some of the ideas underlying standards-based reform. In particular, they do not negate the value of moving toward a system that supports high expectations for *all* students. But the findings do imply that the current heavy emphasis on external testing and accountability—and correspondingly less attention to curriculum, instruction, and professional development—will prevent the ultimate goals from being realized.

We found that districts can have standards, assessments, and accountability in place, yet not improve the quality of curriculum and instruction in classrooms. In contrast, districts that communicate ambitious expectations for instruction, supported by a strong professional development system, are able to make significant changes in classroom practices.

We conclude that clear expectations for instruction are as critical as clear expectations for student learning. Dedicating resources to building the knowledge and skills of educators and to providing additional instructional time for low-performing students is essential if the benefits of standards-based reform are to be realized in increased student achievement.

Creating an infrastructure to support teacher learning, building a rich and constructive assessment and accountability system, and finding the right balance of authority and discretion present significant challenges to district leaders. The fact that few of the Pew Network districts made substantial progress is testimony to the difficulty of the task. Still, we saw signs that educators were taking reform seriously everywhere and, across all the districts, observed several trends over the four years of the evaluation that bode well for continued progress. These trends include:

- Substantially more attention to professional development for teachers, including placement of staff developers in schools.
- A shift away from focusing on a few schools toward including all schools in reform efforts.
- Expansion of assessment systems to incorporate some form of testing that requires demonstration of work beyond checking one of several choices.
- An increase in attention to and use of classroom-based diagnostic assessment with the primary purpose of informing instruction.
- Greater emphasis on supporting principals to become instructional leaders.
- Increased attention to data in school planning.
- Examples of richer notions of accountability that rely on multiple measures, professional judgment, and shared responsibility for student learning.
- More attention to the importance of a district strategy for change that focuses attention on one or two subject areas.
- More opportunities for students who are failing or who are at risk of failure to have extra instruction that is challenging, not remedial.

These trends are fragile, but they have the potential to continue and even accelerate if investments are forthcoming in creating the human capital necessary for reform.

At the same time, extra efforts are needed to help students who start their formal schooling with weak backgrounds or who fall behind at various points along the way—conditions disproportionately true of children of poverty and of color. A commitment to improving instruction for *all* students inevitably means concentrating more resources in the schools and classrooms with students who are farthest from reaching the standards. Here is where the rhetoric of standards-based reform runs most directly into entrenched beliefs and interests in maintaining the status quo, including the placement of less-qualified and less-experienced teachers in the lowest-performing schools.

The bottom line is that taking standards-based reform seriously has profound implications for the priorities, organization, and resource allocation of school districts. Difficult as it is to create ambitious standards and rich assessments, it is far more difficult for districts to equip central office and school staff with the knowledge and skills needed for their new roles. Yet

without these changes, the promise of standards-based reform to increase learning for all students will not be realized. Across the Pew Network districts, the greatest strides occur where the adults also have opportunities to learn.

CONTENTS

Executive Summary ..... **i**

I. Introduction **1**

Background and Theory 1

The Districts 4

The Evaluation 4

Overview of Report 6

II. Standards, Assessment, and Accountability **9**

The Role of Standards 9

Changing Assessments 11

Accountability in Practice 13

Promising Directions 15

Conclusions 18

III. Strategies to Support Improved Teaching and Learning **19**

Expectations for Instructional Practice 20

Strategies to Change Instructional Practice 23

Strategies to Help Low-Performing Schools and Students 28

Conclusions 30

IV. New Roles and Relationships for Districts **31**

How Districts Organize to Support Professional Development 31

School Autonomy and Flexibility 33

The Role of Unions and the Community in Supporting Reform 36

Conclusions 38

V. Conclusions **41**

Building an Infrastructure for Changes in Classrooms 43

Creating Assessment and Accountability Systems That Motivate Improvement 44

Balancing Mandates and Discretion 45

Promising Trends 45

Figures

Figure 1: The Theory of Standards-Based Reform 3

Figure 2: The First Link in the Logic of Standards-Based Reform 9

Figure 3: The Next Links in the Logic of Standards-Based Reform 19

Figure 4: The Original and Modified Theories of Standards-Based Reform 42

References **49**

Appendix: Characteristics of Pew Network Districts

Table A-1: Size of Pew Network Districts A-1

Table A-2: Demographics of Pew Network Districts A-1

I. INTRODUCTION

In 1996, The Pew Charitable Trusts gave four-year grants to seven urban districts to assist their efforts in implementing standards-based systemic reform. Called the Pew Network for Standards-Based Reform, the seven districts are: Christina, Delaware; Community District 2, New York City; Fayette County, Kentucky; Pittsburgh, Pennsylvania; Portland, Oregon; San Diego, California; and Yonkers, New York. The Trusts' primary goal in funding these districts was to test the theory and assumptions behind a standards-based systemic approach to reform. Several times each year, an evaluation team conducted interviews in the districts to track progress in their reform efforts in order to inform a set of broader questions: What does the theory look like in practice? Do the presumed relationships hold up? Do they result in higher achievement for all students? This fourth and final evaluation report synthesizes findings across all four years of the Pew Network.

Background and Theory

The Pew Charitable Trusts both contributed to and anticipated the national importance of standards-based reform. In designing the Pew Network initiative, the Trusts combined the ideas of systemic reform with those of performance standards and assessments embodied in the New Standards Project. These ideas were developed in response to well-documented problems with previous education reforms, including piecemeal strategies that result in different parts of the system pushing in different directions; a focus on inputs instead of results; and low expectations, especially for poor students and students of color (Smith & O'Day, 1990; National Center on Education and the Economy, 1990). In the context of an increasingly global economy and international comparisons of school performance, standards-based reform has been hailed as the way to strengthen public education for all students.

Standards-based systemic reform is premised on the argument that increasing student achievement requires: (a) a clear and ambitious set of expectations for what students need to know and be able to do (standards), (b) assessments and accountability aligned with the standards, and (c) sufficient professional development and school autonomy to implement curriculum and instruction needed to help all students reach high standards.

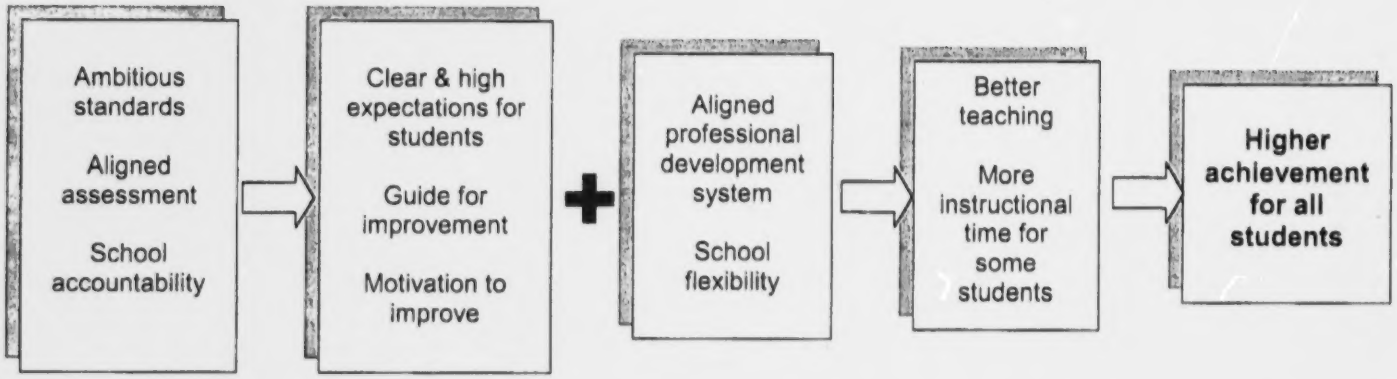
Standards are intended to set learning goals for students that are ambitious and clearly understandable by teachers, students, and parents. Content standards, sometimes embedded in curriculum frameworks, spell out the skills and concepts to be taught; performance standards communicate what students must be able to do to meet defined content standards. Assessments

that adequately measure attainment of the standards are to provide the basis for tracking progress and guiding improvement. Such assessments are expected to differ from traditional norm-referenced multiple-choice tests in two ways: by referencing standards instead of norms and by measuring more complex skills, often in the form of open-ended or performance items. Because it is generally accepted that "what is tested is what is taught," more complex and demanding assessments are expected to be a constructive influence on instruction.

Holding educators accountable for improvement based on the assessments is intended to motivate changes in practice. In exchange for accountability, schools have the flexibility to determine how best to increase student performance. Professional development in support of teaching to standards is the primary process by which the structural elements are translated into changes in practice (Elmore & Rothman, 1999). Coupled with more learning time for those students who need it, these elements are intended to motivate and prepare teachers to implement instructional practices that will enable all students to reach high standards. Together these pieces represent the logic of standards-based reform (see Figure 1).[1]

Translating the theory into action is another story. Although this logic may seem straightforward, a shift from norms to high standards for all students presumes a very different set of beliefs and roles than currently exists in most districts. Promoting teaching practices designed to help *all* students reach ambitious standards runs counter to widely shared beliefs about the nature of learning and about the abilities of many students, especially poor and minority students. Implementing the ideas is further complicated by the fact that standards-based reform is enacted at both the state and district levels, often in different ways. States have become major players in setting standards, assessments, and accountability, but their instructional guidance is not necessarily coherent or consistent with district directions. In this light, creating a standards-based system presents a daunting, transformational challenge for districts.

Figure 1
The Theory of Standards-Based Reform



The Districts

The Trusts targeted midsized urban districts for the Pew Network and selected five such districts and two subdistricts of large systems. Five of the seven districts have student populations between 20,000 and 40,000; Portland is somewhat larger, with roughly 54,000 students. San Diego is much larger, with more than 142,000 students, but began its participation in the Network represented by one cluster with only 16,000 students. In the third year of the Network, San Diego's new administration reconfigured the district's clusters and expanded its reforms to include the entire district. As part of the larger New York City Public Schools, Community District 2 then remained the only subdistrict in the group.

The districts' white student populations range from 21% in Yonkers to 75% in Fayette. District 2 has the only substantial Asian population (32%), and three districts have significant Hispanic populations: Yonkers (43%), San Diego (37%), and District 2 (20%). Pittsburgh has the largest African-American student population (57%), followed by Christina and Yonkers (30% each). The percentage of students on free and reduced-price lunch, as a proxy for family income, ranges from 32% in Christina and Fayette to 66% in Yonkers. Tables A-1 and A-2 in the Appendix provide these data for each district.

Beyond demographic data, each district represents a unique combination of local and state contexts and leadership, all of which shifted during the four years of the Network's existence. As is the case with most urban districts, the Network districts had considerable changes in leadership. No site has had the same superintendent since the network began; four have had more than two superintendents during this time, and two are on their fourth.

Pittsburgh and Portland faced successive years of large budget deficits. Yonkers grappled with its continuing desegregation court case. Both Fayette County and Christina are in states that were well into the development of standards and assessments. San Diego saw a major shift in state direction. District 2 joined the Network several years into its reform and became the subject of a major federal research effort when the Network began.

The Evaluation

The purpose of the evaluation was to view the sites as testbeds for the power and feasibility of standards-based reform—does the theory pan out in practice? In reality, of course, districts do not operate by theories. Theories are general and do not carry blueprints for enacting them, and they rest on assumptions that may or may not hold up in practice. Moreover, they are interpreted

differently by different players, some of whom may not even subscribe to the ideas, and they are carried out on the shifting sands of state and local politics.

During the four years in which we studied these districts, several major shifts occurred that belied the expectation of steady progress in implementing reform. As we described above, we witnessed considerable turnover of top leadership, which typically resulted in new directions for reform efforts. At the same time, each state has become more active in standards-based reform, particularly in the realm of high-stakes assessment and accountability.

Against this backdrop, it is perhaps not surprising that we found early on that standards-based reform agendas in districts are often not readily identifiable strategies; moreover, to the extent that strategies are identifiable, they are likely to change. Hence, our investigation of how districts interpret and enact standards-based reform required continually ascertaining their theories of action or intentions and tailoring questions to the particulars of each site, rather than measuring their actions against a preconceived set of reform features.

Our evaluation followed a modified multiple-case-study design. We essentially conducted seven parallel case studies, collecting data in each site under a common set of categories but capturing the specifics of each site's particular agenda and context. Each year, a team of two site visitors made several trips to each site. We followed at least four schools in each site, typically two elementary schools, one middle school, and one high school. We asked district leaders to recommend schools they considered indicative of the impact of the district's reform efforts.

The bulk of our data collection activities were on-site interviews, mostly with teachers and administrators. We also interviewed school board members, community leaders, and association or union leaders. We augmented these data with observations of selected events, such as professional development sessions and school board meetings, and informal observations of classroom activities. Evaluation team members also participated in Network activities, including the annual Fall Institute and Network Council meetings. We then analyzed the data across the sites to identify common themes and their variations in how districts implemented standards-based reform.

We began with a focus at the district level and each year moved successively closer to the classroom, although when leadership and reform agendas shifted, our focus returned to the district level. In the third year, we focused particularly on assessment, and in that year's report, we describe the difficulty of drawing inferences about student achievement (Shields, David, Humphrey, & Young, 1999). Hence, we make only passing reference to achievement scores in this report and refer the reader to the earlier report for our student achievement findings. In this

report and throughout the evaluation, our primary focus has been on the factors closest to student achievement: teaching practices and the ways in which standards-based reform efforts do or do not influence what teachers do and therefore what students learn.

Overview of Report

Individuals in each district have made heroic efforts to pursue elements of standards-based reform. Yet, each district also has encountered powerful countervailing forces, as noted above. Although they complicate the job of theory testing, these forces are the reality of urban school districts. As such, they cannot be treated as noise, but instead must be treated as factors that reformers must grapple with in their efforts to improve urban education.

In our study, we have observed a range of efforts to develop standards, to put in place new assessment and accountability systems, to implement more challenging curricula, and to strengthen teaching and learning. All sites have standards in place, and all have struggled with developing appropriate assessment and accountability systems. Still, we have encountered few examples of systems that have put multiple aligned elements in place and kept them in place over several years, particularly systemic efforts to strengthen instructional practices directly.

These findings demonstrate that the ideas embodied in standards-based reform are exceedingly difficult to realize in urban districts. They do not negate the overall idea of standards—in fact, the exceptions in our sites can be interpreted as support for some of the ideas underlying standards-based reform. In particular, they do not negate the value of moving toward a system that supports high expectations for *all* students. But the findings do imply that the current heavy emphasis on external testing and accountability—and correspondingly less attention to curriculum, instruction, and professional development—will prevent the ultimate goals from being realized.

We found that districts can have standards, assessments, and accountability in place, yet not improve the quality of curriculum and instruction in classrooms. In contrast, districts that communicate ambitious expectations for instruction, supported by a strong professional development system, are able to make significant changes in classroom practices.

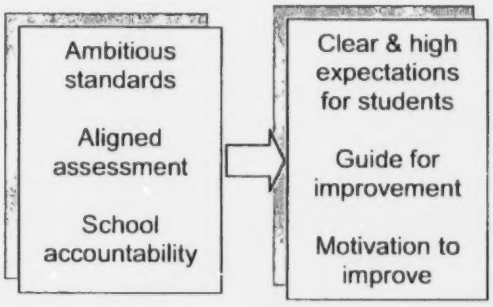
We conclude that clear expectations for instruction are as critical as clear expectations for student learning. Dedicating resources to building the knowledge and skills of educators, as well as to providing additional instructional time for low-performing students, is essential if the benefits of standards-based reform are to be realized in increased student achievement.

In the next chapter, we describe the struggles that districts experience in attempting to put standards, assessment, and accountability in place and the extent to which these elements guide improvement. We conclude that even when teachers are motivated to change, they need more direction and education than standards and assessments provide. In Chapter III, we describe district attempts to strengthen teaching and learning through instructional guidance and professional development. Here we also look at the role of school flexibility in the context of teacher learning and finally at district efforts to assist low-performing schools and students. In Chapter IV, we describe some of the new roles and relationships inherent in implementing standards-based reform and how districts have fared on this front. Chapter V presents our conclusions and modifications of the theory of standards-based reform suggested by our findings.

II. STANDARDS, ASSESSMENT, AND ACCOUNTABILITY

In this chapter, we look at the first link in the chain of the theory underlying standards-based systemic reform: that ambitious and aligned standards and assessments coupled with accountability will lead to clear expectations for students, provide information that can guide school improvement, and motivate educators to improve through rewards or sanctions tied to student performance (Figure 2).

Figure 2
The First Link in the Logic of Standards-Based Reform



The Role of Standards

Every district now operates under a set of standards, in contrast to the beginning of the Network, when several had yet to develop or adopt standards. Over the course of the four years, the vocabulary of standards has infused conversations about district reform efforts. In many schools and classrooms, standards appear in posters on the walls. District leaders have led activities to help teachers and parents understand what standards mean. Fayette County, for example, has created a Standards Review Process through which teams assess a school's implementation of standards-based instruction. Others, notably San Diego and Portland, have provided opportunities for teachers to discuss and score student work with rubrics derived from their standards and, more recently, to observe and discuss videos of ambitious teaching—that is, teaching practices designed to help students reach ambitious standards.

Although standards appear to carry considerable rhetorical and symbolic value, they do not serve well to communicate high expectations for students or to guide instruction. When teachers

see benchmark or anchor papers that illustrate different levels of performance, expectations are clearer. And when teachers have opportunities to meet and discuss the standards, their meaning can become clearer still. But such opportunities were rare in the districts, in part because producing examples of acceptable student work for each standard—or group of standards—for each subject for each grade level is an enormous task.

More commonly, states and districts try to clarify the meaning of standards by providing more detailed definitions. Delaware has developed frameworks with specific performance indicators for each standard—for example, noting that students should be able to "estimate, measure and compute the perimeter of polygons" at the fourth- and fifth-grade levels to meet the standard on estimation, computation, and measurement. Pittsburgh, San Diego, and Kentucky have taken a similar approach.

The limitations on the power of standards to communicate clear expectations and provide guidance also arise from the political context in which the standards are developed. In theory, standards communicate high expectations by focusing attention on the big ideas in each subject and on a parsimonious set of learning goals for each big idea. In practice, the result is rarely a concise set of standards focused on a limited number of key concepts. Pressures to enumerate skills as well as, or in place of, fundamental concepts, and to do so for each grade level, result in documents that cover considerable territory. To some, focusing on a few key concepts signifies rigor; to others, rigor demands enumerated skills and objectives.

In addition, districts typically operate under more than one set of standards, any of which can be significantly altered from one election cycle to the next. The Pew Network sites are all in states with state standards, and most sites have adopted their own, as well. Only Fayette and Christina do not have their own sets of standards, relying solely on their states' standards. District 2 uses the New Standards, and Yonkers has developed standards based on the New Standards and the New York State Standards. Portland and San Diego each went through a lengthy process to develop district standards, but teachers and principals must still pay attention to their respective state standards. In each case, local leaders argue that they need to have their own standards to create local commitment to and understanding of standards. In addition, many teachers look to their national associations as a source of guidance, yet these standards may be quite different from state or local standards.

The four-year history of standards and assessments in San Diego illustrates the difficulty of creating and measuring a set of ambitious standards. In the fall of 1996, the San Diego Board adopted content and performance standards, developed over a two-year period and modeled on the standards of national subject-matter associations and the New Standards. At the same time,

the district embarked on developing a standards-based assessment system, including portfolios and standards-based report cards. Two months after the adoption of the standards, a new board reversed this decision and directed district staff to develop standards more focused on skills and for every grade level instead of grade-level spans. The board also put in place an accountability system that had little relationship to the standards.

During this same period, California abandoned its nationally touted curriculum frameworks and two-year-old performance assessment system. The state developed grade-level-specific state content standards, again reflecting an increased focus on skills and recall. The state then adopted a norm-referenced test (SAT-9) and mandated its use in grades 2-11 to serve as a basis for school accountability. San Diego in the meantime has dropped its accountability system so as not to conflict with the state's and has abandoned portfolios as part of its formal assessment; current plans, however, include continued development of standards-based report cards. Most recently, as more and more state funds are contingent on district adoption of state standards (even though such adoption is technically voluntary), San Diego will once again revise its standards by starting with the state's and adding its own.

Changing Assessments

Standards-based reform assumes that new assessments aligned to standards and able to measure more complex skills would be used as the basis for school accountability; and, because measurement draws attention to what is important, these new assessments would be a constructive influence on curriculum and instruction. In fact, when the Trusts launched the Network, they hoped that the New Standards Reference Exam (NSRE), designed to embody ambitious standards and performance assessment, would be used by all the sites. This plan proved infeasible for a host of reasons, including costs, local preferences, and existing local and state assessments.

Nevertheless, all the districts attempted to incorporate new types of tests, either replacing or augmenting traditional norm-referenced multiple-choice tests. As of school year 1999-2000, every district used some measures designed to assess student performance relative to standards and, at a minimum, had open-response questions and other opportunities for students to write or otherwise display their work. San Diego and Yonkers used the NSRE in a subset of schools for the first two years. Pittsburgh and District 2 continue to use the NSRE in all schools. State tests in New York, Kentucky, Oregon, and Delaware are intended to reflect their standards and include performance and/or open-ended items. California will be adding a writing assessment.

District and state efforts to mount assessment systems that go well beyond traditional normed multiple-choice tests have run into both technical and political problems. Public pressure for assessments that report national norms, political pressures to test every grade level, the difficulty of crafting reliable measures not subject to biases in scoring, and the sheer amount of work involved in scoring written responses collectively have slowed efforts to create assessments that measure and report on students' proficiency relative to standards. One result, across the Pew sites, has been frequent changes in assessment systems and an overall movement away from the use of items—most notably, performance items—tied directly to more ambitious standards. Kentucky dropped the performance tasks from its test because of reliability concerns and added a norm-referenced test. In Oregon, the state had to modify an ambitious plan to collect student work samples from all students in three grade levels demonstrating their attainment of the state standards. Delaware dropped its initial plan for a fully standards-based instrument. San Diego is planning to drop its portfolios from the accountability system as it introduces a standards-based report card system; managing and scoring the portfolios is too resource intensive. California dropped from its accountability system the "augmented" items intended to create a better match between the SAT-9 and the state's standards, because they were too difficult. And New York and California have been plagued with problems with test administration, complaints about the fairness of the items, and mistakes in reporting results.

Norm-referenced tests continue to play an important role in most sites and seem unlikely to go away. The entire assessment and accountability system in California is built on a single test, the SAT-9, and San Diego is required to participate in that system. Norm-referenced tests make up a portion of the assessment systems in Fayette, Yonkers, Christina, and District 2. The exceptions are Portland, which continues to use its multiple-choice criterion-referenced test (PALT), and Pittsburgh, which uses only the open-response items from nationally normed tests.[2]

Pittsburgh has put considerable effort into developing an assessment system that is entirely standards based, even though it will continue to administer portions of a nationally normed test. In each grade, the district assesses reading, writing, and mathematics. In grades 4, 8, and 10, it administers the NSRE and a home-grown clone of that exam in math in grades 3, 6, and 9. In grades 5, 8, and 11, it administers the state's standards-based reading assessment. In grades 2, 3, 6, and 9, it administers the SAT-9. However, it administers only the open-ended portions of the SAT-9 with an abbreviated set of reading comprehension multiple-choice questions, and has negotiated with the publisher a customized reporting format that blends the two scores and reports them against Pittsburgh's standards.

What is measured by the assessment often becomes the de facto definition of the learning intended by standards. In Kentucky, the reporting categories on the assessment—novice through distinguished—define the standards. In Fayette County, schools devote considerable energy to aligning what they teach to what the state says will be on the assessment. Educators in Portland spent two years developing standards that built on the state's Core Curriculum Goals and Content Standards, New Standards, and the earlier Portland curriculum frameworks. However, for students to earn a Certificate of Initial Mastery at graduation, they must pass the state tests in multiple subject areas. As a result, Portland teachers focus their attention on the tests rather than the standards. And, as in Kentucky, it is the released and practice items for these assessments that provide practical guidance to teachers, not the standards. In addition, the importance of the tests is magnified by the district's use of test scores as a key yardstick for judging school improvement plans, to which principals' jobs are tied.

Whether it is through the content expected to be on the test, released examples of acceptable student responses, or how cut-offs are defined, assessments signal what is meant by the standards. However, frequent changes in the assessment systems and continued reliance on normed and multiple-choice exams mean that teachers, principals, and the public face competing visions of what is important for students to know and be able to do, standards notwithstanding. As we describe below, these problems are compounded by the dominance of test-based accountability systems.

Accountability in Practice

During the four years of the Pew Network, there has been increasing political pressure in every state and district to create accountability mechanisms. The presumption is that holding schools accountable for results on aligned assessments will motivate teachers to strengthen their teaching practices. In actuality, as we have described above, assessments change often and are not fully aligned with standards. In the best of cases, accountability systems motivate teachers to make some changes in their practice (for example, having students write more often). More often, accountability pressures motivate them to focus on test preparation to avoid sanctions, either for themselves or for their students.

In the Pew Network sites, accountability systems are predominantly test based and are highly political. Coupled with shifting assessments and sometimes unpopular or unworkable consequences, districts find themselves either changing—or being the recipients of state changes in—the way accountability is defined. Kentucky, unusual in its emphasis on continuous improvement for all schools, was forced to retreat from applying sanctions to high-performing

schools whose performance had declined and to put its accountability system on hiatus with the introduction of a new test. Christina dropped its practice of holding principals accountable for standardized test results when the state introduced the new test; ways to create a politically acceptable and feasible new accountability system have not yet been found. San Diego implemented a new accountability plan, expanded it, and then dropped it in the face of a new state system. Portland instituted a series of policies to hold schools and principals accountable that have been implemented very unevenly.

Moreover, we have seen an increase in test-based accountability for students, either through graduation requirements or through promotion from one grade to the next. Several states either have or plan to have graduation exams for students, including California, New York, Oregon, and Pennsylvania.

Even though accountability schemes rarely pan out as envisioned, they do have effects on instruction, though not necessarily those intended. When there are consequences associated with tests (such as ranking schools in the newspaper, threats of reconstitution, or consequences for students), teachers work hard to improve their students' performance on the test. They do this in several ways that can have either positive or negative effects on instruction. One way is to align the curriculum to cover the content that the test is based on. Another is to incorporate activities that reflect the types of items on the tests. A third is to emphasize areas to be tested over other areas—and to emphasize those for which last year's scores were low. Yet another is to stop regular instruction to prepare students for the test by teaching test-taking skills and drilling students on the content and types of items expected to be on the test.

Under the best of circumstances—where standards and assessment remain relatively stable, are aligned with one another, and include opportunities for students to demonstrate mastery of deeper concepts—there are examples of broad changes in teaching practice. In Fayette, the introduction of the state assessment and accountability system has resulted in increases in student writing, more oral and written explanations in response to questions, and more problems that require the application of skills. Similarly, in Pittsburgh, the assessment has played a role in the implementation of *Everyday Math*. As one local educator put it, "All the momentum behind *Everyday Math* in Pittsburgh would not have happened if the New Standards Reference Exam had not been available as an accountability exam. Principals asked for similar exams at other grades."

In a Christina middle school we visited, standards and assessments had focused teachers' attention on student work. Because of the amount of writing on the state assessment, the school staff decided to make every teacher—including the computer lab director and the music and art

teachers—a teacher of reading and writing. The science teachers now require writing in response to prompts, and the mathematics teachers chose to adopt the *Connected Mathematics Program* in part because "we recognize that the test is going to make our kids write explanations for their answers, and we hope CMP will help with that." Across the subject areas, teachers are using rubrics modeled on those in the state assessment to score student work.

When the tests are focused predominantly on skills in multiple-choice format, the response is usually to devote substantial time to test preparation. However, direct test preparation efforts do not embody the kind of teaching intended by ambitious standards, and they often supplant and break the flow of the normal curriculum, thereby diverting teachers from the deeper changes intended by the reform. As one teacher described it, "The workbook modeled on the test can become the literacy program."

The lower the performance of the students, the more time teachers are likely to devote to test preparation. Hence, the very students in need of more exposure to the curriculum can end up receiving less. Even in districts where test performance is downplayed by district leadership, educators whose schools are at risk of sanctions put in extra time to prepare students for the tests. Moreover, when the scores determine placement for students, teachers believe they are cheating their students by not preparing them for the external exams. District 2's after-school and extended-year programs incorporate preparation for the citywide norm-referenced test.

Promising Directions

Standards-based reform in principle assumes that assessment and accountability will lead to better teaching and learning, not simply to increased test scores. The emphasis on test-based results, however, does not appear to lead to serious appraisal of instructional practices. We observed three ways in which assessment and accountability have the potential to improve curriculum and instruction. One is to focus accountability on instructional practices directly. A second is to use data for school planning. A third is to use classroom-based assessments that provide useful information to teachers.

Monitoring Instruction

Accountability systems usually rely on annual test score data or school plans as rough proxies for student learning. An alternative, however, is to look more directly at classroom practice. District 2 leaders regularly monitor what goes on in classrooms; San Diego is putting a similar process in place. Both districts, however, must contend with a state accountability system that relies solely on test scores. As we described in our third-year report:

> [District 2's] now famous "walkthrough" is designed around explicit expectations for literacy instruction. District leaders regularly visit each school and go to every classroom with the principal where they talk to students, look at their work, observe teachers, and look at their data on student growth. They follow up the visit with a letter to the principal that includes changes they expect to see during the next visit.
>
> This inspection has three unique qualities, compared to other measurement and accountability systems: (1) It is grounded in an explicit vision of what good classroom practice looks like—what teachers and students should be doing; (2) it focuses on the lowest performing students (regardless of social class or ethnicity or special education status); and (3) it is based on actual observations of practice and of student learning (including their work and their answers to questions), not a test score that acts as a proxy for such. The "teeth" in this approach lies in the fact that it is the basis for principal evaluations and in turn teacher evaluations. The fact that District 2's test scores continue to increase, while the percentage of students in the bottom quartile decreases, lends validity to their particular vision of quality instruction.... But it is the actual classroom practices and direct investigation of student learning, not test scores, for which staff are held accountable. In contrast to test-based accountability, the rise in scores is not an end in itself but a by-product of focusing on the improvement of practice and student learning. (Shields et al., 1999)

School Planning

Holding schools accountable for test scores increases attention to the scores. Coupled with required school plans, we have seen an increase in the use of data as part of the school planning process. In Fayette County, schools pay close attention to state assessment data in preparing their Consolidated Plan. Weak areas are typically targeted for more attention each year. Kentucky's assessment, as well as Delaware's and the New Standards Reference Exam, provides detailed results that help school planners make decisions about improvement efforts.

Under the best of circumstances, when results point to specific skill deficiencies and arrive in a timely manner, we have seen test results used as a general guide to school planning—for example, looking at patterns of student responses in one school in Fayette suggested the need for more focus on phonics in the early grades. However, as we described in more detail in the third-year report, annual assessment results provide limited guidance, for a host of reasons. They often reflect only one grade level per school. They cover only a small portion of the standards and cannot fully capture the cognitive complexity of demanding standards—what students do and do not understand. Moreover, planners rarely have the knowledge to interpret the data and

devise appropriate remedies. As a consequence, school plans typically reflect short-term strategies, usually increasing the time allocated to the subjects with the lowest scores. These limitations mean that such data are rarely the basis for serious reflection about teaching and how it might change. Nevertheless, paying attention to data and noting relative strengths and weaknesses each year are sensible activities that were not standard practice in many schools in years past.

Assessments to Inform Instruction

The kinds of data teachers need to tailor instruction to individual students' needs are quite different from the gross school-level averages used for ranking schools and even from the finer-grained classroom and student results that represent one point in time. To be instructionally useful, assessment data need to be directly tied to what teachers are doing and provide frequent feedback for each student.

Across the sites, we have seen increased efforts to adopt or develop assessments that can serve these purposes. We have seen a marked increase in classroom-based assessments, especially in early reading. Diagnostic tests, such as the Diagnostic Reading Assessment (DRA); running records, in which teachers track individual students' reading progress at regular intervals; and student portfolios appear in almost every site. In some cases, this use of classroom-based assessment had been going on for a while—such as running records in District 2 and portfolios in San Diego. In general, such assessment is used much more than in the past; for example, in Christina, Pittsburgh, San Diego, and Yonkers, every early-grade teacher is now using a diagnostic reading test to track students.

Typically, these assessments are intended primarily to provide teachers with specific information on a student's progress. For example, the DRA will inform a teacher that a student in the middle of first grade is struggling with initial consonant blends. Unlike a test score that arrives in the fall informing the teacher that a student is in the 49th percentile or in the "novice" category, such information from the DRA can be used to target instruction that very day.

In a growing number of cases, however, the usefulness of these classroom-based diagnostic assessments is compromised by their incorporation into a district's accountability system. Thus, for example, the portfolio process in San Diego—originally designed as a way for both teachers and students to track progress relative to standards by including examples of work that represented a student's development—began to be seen as a compliance exercise when the district started publishing schools' portfolio scores. Similarly, in Fayette County, the portfolios required by the state assessment system are not viewed as instructionally useful by teachers.

Conclusions

Over the four years of the grant, the Network sites have worked hard to put standards and assessments in place, and have attempted to create sustainable accountability systems. But, for both technical and political reasons, one result is the existence of multiple sets of standards, some more ambitious than others, and changing assessment systems that cannot break away from norm-referenced tests. And these are the basis for accountability systems that often are changing, as well.

Nevertheless, all the sites have managed to incorporate some form of testing that requires demonstration of work beyond checking one of several choices. We also see more attention to assessment data as schools develop improvement plans. In addition, we observed a marked increase in the use of classroom-based diagnostic assessments, although their value for instructional improvement was far greater when they were not used for external assessment and accountability purposes. And in two sites, a very different conception of accountability is evident—one that rests on ongoing professional judgment about the adequacy of instructional practice rather than annual test scores.

Overall, the core components of standards-based reform—standards, assessments, and accountability—do not play their intended roles well. They do not do a very good job of communicating high expectations for students, providing information to guide instructional improvement, or motivating widespread instructional change beyond test preparation.

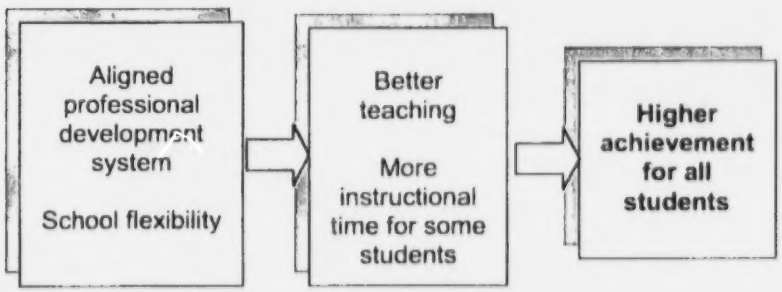
High-stakes accountability does motivate educators to avoid sanctions. However, the goal becomes one of raising test scores, which typically results in less ambitious teaching, especially for low-performing students. Only when the assessments suggest more ambitious teaching, by asking for written arguments or applications of knowledge, did we find teachers attempting changes beyond practicing test-like multiple-choice items.

Even when teachers are motivated to adopt more effective practices, standards and assessments are not sufficient to help them do a better job. Increasing writing activities in the classroom, for example, is not the same as teaching writing better. As the theory suggests, teachers need opportunities and support to learn new practices, which we turn to in the next two chapters. We also look more closely at the sources of ideas and expectations for new practices, given that standards and assessments do not describe what teachers should do or how to do it.

III. STRATEGIES TO SUPPORT IMPROVED TEACHING AND LEARNING

To improve teaching, standards-based reform theory adds two pieces to standards, assessment, and accountability: (1) a professional development system for teachers aligned to standards, and (2) enough flexibility at the school site for faculty to design appropriate educational programs. The final piece is more instructional time for students who need it, which, when coupled with better teaching, is expected to produce higher achievement for all students. (See Figure 3.)

Figure 3
The Next Links in the Logic of Standards-Based Reform



In the preceding chapter, we described considerable work across the Pew sites in the areas of standards, assessment, and accountability, which send some signals to teachers about expectations for student learning. However, even in the few cases when standards and assessments are reasonably well matched and consistent from year to year, at best they serve as a starting point. They do not suggest what teachers should do or how to do it.

In this chapter, we look at what districts do to take the next step: helping teachers learn not just what students should know and be able to do but what they as teachers can do that leads to such student learning. We focus first on how districts communicate expectations about instructional practice to teachers. We then examine district professional development strategies aimed at helping teachers meet these expectations. We pay particular attention to district efforts to support improvement across *all* schools and across *all* classrooms. Finally, we look at the ways in which districts provide assistance to low-performing schools—a form of professional development—and how they provide additional time for low-performing students.

In Chapter IV, we look at how districts reorganize to support widespread changes in teaching practices. There, we discuss the degree to which the flexibility to design new educational approaches at the school level plays a role in improving teaching.

Expectations for Instructional Practice

In the Network sites, we found that teachers who were changing their practices not only got clear messages about what students should know but also had access to a respected source that offered a clear set of expectations for curriculum and instruction. For individual teachers, such a vision can come from peers and professional associations or through professional development and experimenting with a new curriculum. Within a single school, the principal can be the source. Widespread changes beyond a few individual classrooms and schools, however, can be traced to a clear and consistent message from the central office about their expectations for curriculum and instruction.

We observed several ways in which districts communicate specific expectations for teachers. The most common way is to adopt or adapt a new curriculum that embodies ambitious standards, which signals teachers about the kind of instruction that is valued. For example, in five of the seven Pew sites, the districts have adopted the *Connected Mathematics Program* (CMP) for middle school mathematics. Designed to be consistent with the NCTM standards, CMP contains a series of units that embody demanding mathematical concepts and require interactions between teachers and students and among students that are quite different from traditional modes of mathematics instruction. For example, instead of being told the formula for the circumference of a circle, small groups of students measure the diameter and circumference of a variety of circles, graph their results, and describe the pattern they find. Then they are told the "shortcut" (formula).

Districts also communicate explicit expectations for instruction through a framework for instructional practice, rather than a single published curriculum. District 2 and San Diego both use a balanced literacy framework—not a traditional textbook adoption, but nevertheless an explicit statement about expectations for literacy instruction—that includes the types of materials to be used, as well as physical arrangements (such as readily accessible materials, including classroom libraries of children's literature labeled by reading level). San Diego's Literacy Framework, for example, calls for teachers to have a three-hour block of literacy instruction each morning that incorporates each of six reading elements: Reading Aloud, Independent Reading, Word Study, Observation and Assessment, Shared Reading, and Guided Reading. Each element is described in more detail; for example:

> Independent reading by students gives them opportunities to practice the strategies they have learned in shared reading, guided reading, read aloud and word study. Teachers provide guidance with book choices, tailor teaching to meet individual needs and meet with individuals to monitor progress. Books from a range of levels are available in the classroom. Students become proficient at selecting books that match their interests and reading level. (San Diego K-12 Literacy Framework; http://www.sandi.net/comm/parents/lit.framework.htm)

In addition, districts signal expectations for instruction through promoting a set of principles that characterize effective classrooms across subject areas. Several of the Network districts are using the University of Pittsburgh's Principles of Learning, often in conjunction with a specific curriculum, to guide teaching practice. These principles capture ideas such as the importance of effort over aptitude, worthwhile classroom talk that is accountable to others, and fair and credible evaluations (http://www.instituteforlearning.org/pol3.html).

Whether districts communicate expectations for curriculum and instruction through a textbook adoption or a new framework, the real challenge is making those expectations a reality in each classroom in the district. The challenge is twofold. First, district leaders must overcome the skepticism of school staff who are wary of adopting new reforms that too often are quickly abandoned. Second, even when teachers and principals agree with and are committed to the new expectations, realizing them in the classroom is very hard work.

Teachers and principals alike view with suspicion any change in signal about curriculum and instruction on the assumption that it will be short lived. And their assumptions are well grounded in experience. In the Network sites, District 2 is the exception, having sustained its balanced literacy framework from several years prior to the Network through today. All the others have changed course one or more times during the four years of the Network's existence.

As districts face pressure to show test score results quickly, attempts to put new ideas in place do not have the luxury of District 2's decade-long evolution. San Diego is attempting to put a similar balanced literacy framework in place over a much shorter time. Yonkers' new superintendent introduced both a literacy framework and a new mathematics curriculum at the same time—requiring their immediate use in all schools. Even District 2 moved more quickly in mathematics, formally adopting a new curriculum in both elementary and middle schools.

When districtwide instructional policies are put in place quickly, the quest for efficiency preempts teacher buy-in. Since most teachers are already trying to do a good job, pressures to change how they teach, how they organize their rooms, and how they assess their students are

often met with resentment and resistance. Moreover, differences in instructional approaches across schools and among teachers within a school are always in tension with the goal of requiring everyone to implement one particular view of effective practices.

Yet, when new practices make sense to teachers, we see them willing to make efforts to put them in place. The problem is that doing so is not easy. Across the schools and classrooms we visited, we were impressed by the difficulties that teachers and principals face in fundamentally changing their practice. Where new and more ambitious practices are expected of teachers, they often need to learn new content as well as new teaching strategies. For example, a sixth-grade teacher who has followed a curriculum of arithmetic is now asked to teach concepts of probability and statistics—ideas that may be as foreign to her as to her students. Similarly, a teacher who is accustomed to reading a story aloud to a first-grade class to introduce them to the joys of reading is now asked to model a set of reading strategies while she reads and to make those strategies explicit to the students. Silent reading time in third grade, when a teacher might catch up on some of his paperwork, now requires the teacher to select a set of leveled books for each student to choose from and monitor students' individual progress. Informal assessment of students' progress is now augmented by regular maintenance of running records and periodic use of formal diagnostic assessments. As one Christina teacher describes it:

> Things have changed radically with the *Four Blocks*—at least the way we are doing it. We are sharing more often. We are taking more ownership—I feel like I have more responsibility, and I have learned that the kids can do much more than I ever thought. Specifically, I spend a lot more time on writing. And I see a difference in the quality of the writing—because I have changed how I teach. For example, in the past, if kids got words on the page, they were spelled correctly, and they could read to me, I was happy. Now I am looking for a strong beginning, a strong middle, and a strong conclusion to the piece, as well as evidence of revision, editing, etc. In reading, we used to use this end-of-unit multiple-choice test, and kids would just guess. But now we do a running record from each child. So this is a big difference; I used to have little data from the end-of-the-unit exams—now, with the running record, I really know what kids can do.

Simply put, teaching more demanding curricula well requires teachers to do much more and to expect much more of their students. These expectations in turn call for many teachers to change their attitudes about what students—especially the lowest achievers—are capable of doing.

Strategies to Change Instructional Practice

Given these challenges, it takes a number of years, support from principals and central office staff, and a rich system of professional development for teachers to move beyond "going through the motions." Creating such a system is a complex undertaking, because districts must reach *all* teachers and principals and meet a wide range of needs.

Across the Pew sites, central office staff are seeking ways to support teachers and principals, often with limited capacity and resources. Where districts are able to provide such support, they do so by providing both teachers and principals with extensive opportunities to become immersed in new ways of teaching content and in the ways students learn it. They provide opportunities to see examples of good practice, to have help in trying out new strategies in their own classrooms and schools, and to have time to talk about and reflect on their practice with colleagues.

Fayette County, for example, has dramatically changed its approach to professional development in the last year. Replacing the former system of menus of workshops and districtwide events, the district has consolidated its professional development and curriculum resources and targeted them to school-based assistance in language arts and mathematics. District curriculum specialists now spend virtually all their time in schools working directly with teachers. Six spend most of their time in three of the lowest-performing schools with the most inexperienced teachers. In addition, eight master teachers have been freed up full-time to work with beginning teachers on reading instruction. Every middle school and high school now has a literacy specialist. The district pays tuition for middle and high school teachers to receive their certification as literacy specialists from the University of Kentucky.

Providing such an array of learning opportunities strains district capacity. Budget cuts, especially in Pittsburgh and Portland; traditions of school autonomy, especially in Christina and Fayette; and shifts in leadership and direction, especially in Yonkers, limit what districts can do. Moreover, reaching all teachers in sufficient depth requires districts to have, find, or train a large cadre of providers of professional development.

In the face of these challenges, districts have taken steps to make the task manageable. Most chose early on in the Network to focus professional development on one or at most two content areas, usually literacy and mathematics, and several limited their focus to elementary schools. Typically, the professional development was provided in and tied to the adoption of a new curriculum. With the exception of District 2 and, later, San Diego, the amount of time available for professional development has been severely limited, in part because of the cost.

External grants are frequently the mainstay of professional development activities for this reason, including in District 2 and San Diego.

Early in the Network, most districts also focused their reforms on a subset of schools, but over time there has been a shift away from this practice. Initially, Christina and Yonkers provided intensive development to a small number of schools. Similarly, Fayette began with a group of schools and trained a new group each summer. The former San Diego Cluster, a subset of San Diego schools, received substantially more professional development from its leadership than most other clusters. All these districts now have districtwide efforts in lieu of targeting a subset of schools.

Districtwide efforts, however, differ in the extent to which they influence all schools and teachers. Reforms in mathematics in Pittsburgh and in literacy in Christina demonstrate how broad change efforts can have real impacts in schools and classrooms—impacts that result in improved student performance—yet still fail to have districtwide results.

Pittsburgh, for example, adopted new mathematics curricula at both the elementary (*Everyday Math*) and middle school levels (the *Connected Mathematics Program*). The central office has a strong mathematics leader, who is assisted by six mathematics demonstration teachers. The demonstration teachers conduct monthly workshops on particular units of the new curricula and are also available on call for assistance in the schools. These new curricula are generally aligned with the New Standards Reference Examination, which gives teachers specific feedback on student progress.

Pittsburgh's progress has had clear impacts on student achievement. Mathematics test scores have risen overall, and the program has reduced the gap between African-American and white students (Briars & Resnick, 2000). Yet, the math reforms are also marked by great variation across schools. The district's own analyses demonstrated that those schools in which most teachers were "high implementers" of the mathematics reforms had significantly higher gains than schools in which most teachers were "low implementers." And, according to district staff, "there are not enough demonstration teachers to go to scale." Consequently, changes in classroom practices typically are a function of the strength of the school principal and the motivation of individual teachers. Entrepreneurial principals who are able to garner additional outside resources and who share the central office's vision of reform are able to create more effective learning environments for their teachers—and better student results—than their less influential peers. Individual teachers who buy into the reform and are motivated to attend district workshops and seek assistance from the limited pool of district staff are more likely to realize more fundamental changes in the classroom. Pittsburgh is now planning to expand this approach

to include more schools and to use the successful elements of the mathematics work in other subject areas.

Christina provides another example of an intensive effort to provide teachers with support for adopting new teaching strategies. As part of its focus on early literacy, Christina has introduced a new balanced literacy curriculum, *Cunningham's Four Blocks*. Along with the new curriculum has come an expanded professional development program. Teachers who choose to participate attend three to four introductory sessions, followed by monthly districtwide grade-level meetings. Teachers are then visited by their staff developers (each is responsible for four elementary schools), who model lessons and coach teachers. New teachers or those who report that they are struggling with the new strategies are likely to be visited once a week by a staff developer; more veteran teachers are much less likely to have a staff developer come to their room. One teacher reported, "The first year, the staff developer was in my room all the time helping me. But now that I am more comfortable with the program, I rarely see her in my classroom—we meet once a week as a team." A number of schools also have lead teachers, who have some release time to work with their peers.

As in Pittsburgh, this approach has had uneven results. In Christina, a tradition of school autonomy has meant that individual schools choose whether to participate in district initiatives, and in some schools individual teachers make choices. A few elementary schools in the district have simply not participated in the *Four Blocks* curriculum, whereas others have embraced it enthusiastically. One school that adopted the new approach was the lowest-scoring elementary school in the state writing assessment three years ago. Most of the primary teachers took advantage of every possible professional development opportunity to learn the new curriculum and, after three years, the school moved from last to third in the state.

Among the Pew sites, only District 2 succeeded in building an intensive and comprehensive professional development system that has been demonstrably effective in terms of both classroom practice and student learning districtwide—although creating such a system began many years before the Pew Network started. Schools and teachers in District 2 do have some discretion—some more than others—but the combination of deeply embedded norms for practice embodied in a districtwide approach to literacy and mathematics and the self-selection of teachers into and out of the district has resulted in a remarkable commonality of practice across all the classrooms. The appearance of each classroom and the kinds of activities that occur are quite similar within and across schools. New teachers quickly become a part of this way of operating. We observed more consistent attempts at ambitious instruction in literacy and math in District 2 classrooms than in those of any other district—and we have seen a positive impact on

student achievement, using both norm-referenced and standards-based assessments, particularly for the lowest-quartile students (Shields et al., 1999; see also Stein, Harwell, & D'Amico, 1999).

The norms and the changed practice in District 2 result from a rich array of learning opportunities for teachers, which include[3]:

- Summer institutes for teachers that provide opportunity for in-depth study of content and how to teach it.
- Staff developers at school sites to support teachers in their classrooms.
- Extended opportunities (three weeks) for teachers to observe master teachers.
- Peer networks and study groups.
- Master teachers assigned to the lowest-performing school.
- Support to participate in an array of university courses designed for teachers.

In its mathematics reforms, now in their fourth year, the district has altered its strategy somewhat, relying more on teacher leaders at each school than on full-time staff developers.

District 2 has treated principals as the linchpin of change and invested heavily in recruiting, training, and supporting its "heads of school." Principals are trained to support teachers and are held accountable for teacher practice. The district sees its role as supporting principals to support teachers. Monthly principals' conferences focus exclusively on instruction.

District 2 built its system over a long time and has had several advantages over other urban districts by virtue of being part of a larger system. Therefore, it has been particularly interesting to watch the former District 2 superintendent attempt to build a similar system in San Diego over a much shorter time. Because this new direction began in the later years of the Network, our observations cover less than two full years. Nevertheless, the impacts of these recent changes in the system have already been felt by every school.

Like District 2, San Diego has focused its resources on professional development for teachers and principals in support of its Literacy Framework. Teachers attend week-long summer institutes in literacy and have access to staff developers at the school site. In their newly created positions, staff developers receive both additional training and assistance from district instructional leaders, who have replaced assistant superintendents. Every school has at least one half-time staff developer, and those with the lower scores have two full-time staff developers.

Also like District 2, San Diego places great emphasis on helping principals become strong instructional leaders in their schools. Visits from district instructional leaders, which include classroom walk-throughs and follow-up letters; requirements to spend two hours each morning in

classrooms; and staff development focused on literacy instruction are all pieces of this effort. Although the new way of doing business is less than two years old, every school has felt its impact strongly, and the language of the Literacy Framework and associated practices can be heard at every level and in every school[4] throughout the district. As we describe in the next chapter, creating such a system requires a major reallocation of resources and central office staffing—and creating it quickly requires nonnegotiable demands from the central office, which are not received favorably by principals and teachers.

Although no other district has been able to mount as comprehensive a system as District 2 has or San Diego is putting in place, many of the Pew sites have begun to focus more attention on strengthening principals as one way of spreading reforms districtwide. In these districts, central office staff meet regularly with principals to communicate information and discuss administrative matters, often pertaining to new regulations or requirements from the state. As these sites have moved forward with their reform agendas, there has been some movement to make principals' gatherings more content focused. Fayette has shifted its principals' meetings from an administrative focus to one on best practices in literacy. Portland's principals' meetings now include presentations from the district's content specialists on what principals should be seeing in their own classrooms in basic literacy and mathematics. In Christina, principals are increasingly taking part in the curriculum workshops offered in mathematics and literacy. In each of these sites, we witness a slow evolution toward building principals' capacity, along with that of teachers, in implementing a new vision of classroom instruction.

As principals are asked to play a more active role in monitoring teacher performance and helping teachers improve, their knowledge of curriculum and instruction becomes increasingly important. However, these new roles are added to an already complex job of managing a school site, often including responsibility for its budget. In every site, educators described these increased pressures. One consequence is difficulty attracting good candidates. All the districts are concerned about attracting strong candidates to principalships, and several have taken steps to develop the leadership potential of teachers. In Fayette County, the district has created a new position at some elementary schools called "principal's assistant," for which teachers in the school apply, many of whom already have their administrative credential but have never intended to use it. This position acts as a stepping-stone for then applying for principalships in other schools. District 2, with its stringent demands on principals, has created its own program for "aspiring principals," in conjunction with a local university. San Diego also is working with a local university to build leadership in the district, and it appears that the role of school staff developer will become a pathway to principalships.

Strategies to Help Low-Performing Schools and Students

Even with clear directions for instructional improvement and a rich professional development system, some students and some schools overall do not achieve at acceptable levels. Whether the results have more to do with students' prior histories of failure or with the quality of teaching, districts (or their states) must respond, for the goal of standards-based reform is not simply to punish those who fail, it is to stimulate conditions that guard against failure and to provide extra time for students who need it. The Pew districts are struggling to devise effective strategies to assist low-performing schools and to provide additional instructional time for students who are not meeting standards.

Kentucky has a strong record of supporting low-performing schools. Its Highly Skilled Educator intervention provides full-time assistance for two years to the lowest-performing schools with declining test scores. Fayette has recently embarked on its own intervention, targeting certain low-performing schools with extra teacher coaching. District 2, in collaboration with the teachers' union, provided two master teachers with enhanced salaries to the school deemed unsatisfactory by the state. And District 2 calls attention to its low-performing students by reporting test scores by quartile for each school. San Diego has assigned two (instead of one) full-time staff developers to the lowest-performing schools and has increased their budgets.

In general, however, the Network districts and their states put more effort into identifying low-performing schools and students than into supporting their improvement. As a result, most interventions provide only minimal assistance and resources. For example, New York State's intervention has been to send in a team to identify needs and require an improvement plan from the school. California calls for low-performing schools, identified broadly as all those below the 50th percentile, to volunteer to receive $50,000 and some consultant time and, like New York, requires an improvement plan. Those schools whose plans are approved receive $168 per student per year to implement their plans over two years.

Among the Pew Network sites, few have mounted significant efforts to provide additional instructional time for students at risk of failure. Although an underlying premise of standards-based reform is the need for some students to have more time, few sites have made a major investment in this area. Kentucky provides funds for Extended School Services, but the design is left up to each school, and the result has typically been traditional after-school remedial programs. Yonkers provides an after-school tutoring program for underperforming students. In addition, the district encouraged elementary schools to deploy all teachers to special small-group instructional periods several times a week to bolster reading and mathematics skills of low-

performing students. Christina, like Yonkers, offers an after-school program, and both have run into transportation issues. In Pittsburgh, schools decide what services to offer.

The two exceptions are District 2 and, more recently, San Diego. District 2 has substantially increased time for low-performing students, in both extended-day and extended-year classes that are small and tightly focused on literacy instruction. San Diego has launched its new Blueprint for Student Success, similar to the strategy of District 2, which incorporates strategies for prevention, intervention, and retention. Prevention includes a double-period literacy class for all sixth-graders, as well as extra staff developers, in schools with the lowest test scores. The intervention strategies, targeted to students at risk of failure, include before- and after-school extended-day classes, double blocks of literacy and mathematics, and summer and intersession programs. In both District 2 and San Diego, the extended programs incorporate the same ambitious instruction as the regular program and are designed to make extremely efficient use of time.

Mounting such programs is expensive, requiring substantial investments not only in additional staff time but also in professional development for the staff. Without investment in professional development, providing extra instructional time for students is often ineffective. Traditional remedial after-school and summer school programs did not have a strong track record in any of the Network districts.

Finally, we heard again and again from teachers that they face problems that are not academic in nature. Students—more and more, according to many of our respondents—bring problems to school that interfere with both their learning and that of others. Teachers and principals expressed the need for resources to cope with these problems, which are compounded by high mobility rates, often for these same students. Dwindling support for school nurses and social workers at a time when health and emotional problems are perceived to be on the rise leaves school staff frustrated by their inability to cope. Much of the stress of accountability is felt around these issues, which many educators believe are beyond their control. When asked what they would do with an additional $100,000 in their budget, several elementary principals in District 2—those viewed as leaders in the district—replied that they would hire a social worker rather than another teacher. In Kentucky, the school-based Family Resource Centers established by the state reform legislation are highly valued, but the demand far exceeds what they can provide. Across the Pew sites, districts simply do not have the resources, staff, or expertise to provide needed social service support to students.

Conclusions

Fundamental classroom changes result when teachers have a clear set of expectations for curriculum and instruction in addition to expectations for student performance. Such changes occur districtwide only when such expectations are backed up by opportunities for teachers to see examples of good practice, to have help in trying out new strategies in their own classrooms and schools, and to have time to talk about and reflect on their practice with colleagues.

Across the districts, we have observed a significant shift in strategies designed to affect instructional practices, from focusing on a small number of schools to targeting all schools. We also have seen an increasing tendency for districts to focus on one or two subject areas, typically literacy and mathematics. Although there are risks in ignoring other subjects, intentionally focusing attention and resources for a period of several years seems essential for changes to occur in classrooms on a wide scale.

The most effective strategies to change practice that we observed involve the creation of professional development opportunities for all teachers and principals, school-based staff developers who work with teachers, and a supervisory system, all grounded in the goals of an explicit vision of effective instruction. Such systems are expensive and require major reallocations of resources—both dollars and people. Hence, implementing such a strategy takes bold leaders who offer a vision of instructional practice that makes sense to teachers and who put significant resources behind realizing that vision. Later, we discuss the downside of strategies that rely heavily on strong district mandates.

We also found that districts, and their states, dedicate disproportionately more resources and attention to identifying failures than to prevention or assistance. Few of the districts, or their states, have made substantial investments in preventing failures from occurring through early intervention. Nor have many invested in intensive strategies to assist failing schools and students.

IV. NEW ROLES AND RELATIONSHIPS FOR DISTRICTS

In this chapter, we describe how the districts reorganized their central offices to provide direction and support to schools. We then turn to the implications of these actions for the relationship between districts and schools, focusing on one remaining piece of the reform theory: autonomy and flexibility for schools to design programs best suited to their particular context and needs. Finally, we examine the implications for district relations with their unions and the broader community.

How Districts Organize to Support Professional Development

Districts that succeed in supporting widespread and ongoing improvement in teaching practice have shifted their central offices from ones that manage dollars, programs, and people to ones focused on leading and supporting improved instruction. This shift carries implications for both the kinds and numbers of positions in the central office and how their roles are defined.

For all districts, such reorganization is extremely difficult. Central offices have typically been built up over the years in response to a variety of special programs (e.g., Title I, special education, magnet programs) and the need to oversee complex functions (e.g., busing, maintenance). As a result, many central offices consist of a set of disconnected fiefdoms, each of which was created to meet a specific purpose but which together do not add up to a comprehensive and coherent structure in support of schools and teachers. Most importantly, the typical central office is not organized to support the instructional mission of the schools.

Making wholesale changes to central offices to provide increased support for instructional improvement requires dismantling some of these fiefdoms, reducing resources in some areas while increasing them in others, and retraining district staff to play new roles—or, in some cases, reassigning staff and hiring new individuals. Such changes are difficult to accomplish in many urban districts because of the entrenched political power of certain offices and the inevitable tensions associated with reassigning staff. Strong leadership and perseverance are necessary to move in this unpopular direction; for most urban districts, the high rate of turnover of superintendents and school board members preclude making such wholesale changes.

As a result, districts are more likely to make some modifications in their organizational structure or invest discretionary funds in certain offices rather than make wholesale changes in the district organization. Christina, for example, began its work in the Pew Network by providing release time for strong teachers in a subset of schools to support their peers. Soon, the district leadership recognized that this strategy was insufficient to produce broad changes within

those schools and across the district. Consequently, they substantially increased their investment in central office professional development services. Direct district expenditures on staff development have risen from around $200,000 at the beginning of the Pew initiative to over $1 million. There are now 12 district-level staff developers, up from 2 at the beginning of the initiative, who spend virtually all of their time working with classroom teachers—both running workshops and coaching in classrooms. The staff developers are supported by two district supervisors, one in math and one in language arts.

Pittsburgh took a further step and reorganized that part of its central office responsible for instructionally related issues. The district consolidated the offices for curriculum, instruction, assessment, and professional development under one Chief Academic Officer. The reorganization is intended to signal the importance of linking curriculum, instruction, assessment, and professional development. Fayette County is in the process of a significant reorganization and redefinition of central office functions that influence instruction most directly. The proposed restructuring moves many of the functions related to student achievement under one part of the organization, including professional development and research and evaluation. These offices were previously in different parts of the organization. Plans also include shifting the roles of many central office staff to spend much of their time in the schools. Under a new superintendent, Yonkers also reorganized its central office and concentrated resources behind the superintendent's reading program.

San Diego provides an example of an effort to fundamentally restructure the entire central office in support of instruction. The San Diego strategy has been to slowly build a new organizational entity—the Institute for Learning—inside the old central office. As new roles are created and filled inside the Institute, external positions that do not fit well are eliminated. The major components of the new structure include:

- The creation of a new top-level post, Chancellor of Instruction, reporting directly to the Superintendent and responsible for all instructional decisions in the district.
- The creation of the Institute for Learning, headed by the Chancellor of Instruction, housing all district programs focused on teaching, learning, curriculum, assessment, and professional development.
- Consolidation of all functions relating to standards, assessment, accountability, and program evaluation into a single office in the Institute for Learning.
- Elimination of the position of Assistant Superintendent and creation of the position of Instructional Leader in its place. Each of seven Instructional Leaders is responsible for

overseeing all instructional activities in 25 of the district schools. Principals report directly to the Instructional Leaders.

- Elimination of 104 additional positions in the central office, resulting in savings of $8.4 million, all of which was then allocated to support school-based staff developers for every school.
- Redirection of Title I funds from teaching assistants to staff developers and other supports for struggling students.

The rationale for the San Diego changes has been to concentrate resources to directly support learning of new instructional practices by teachers and principals. Modeled in some ways after the organization of the District 2 central office, such a radical restructuring is rare in urban systems. As in District 2, San Diego's approach has required a significant reallocation of resources, particularly to support the staff developer positions in the schools. Also, like District 2, San Diego has had to turn to outside funding to launch an initiative in mathematics similar to that in literacy. San Diego, however, is considerably larger and, unlike District 2, houses all district functions, which in New York are split between community districts and the New York City central office.

These major changes in roles and budget allocations have had a dramatic impact on schools in San Diego, in terms of both their discretionary resources and their historical decision-making authority. In fact, across the Pew sites, we found that the stronger the district role in guiding instruction, the greater the resulting tensions between the central office and the schools, on the one hand, and unions and the community, on the other. We next turn to this issue: the role of school autonomy and flexibility in the implementation of standards-based reform.

School Autonomy and Flexibility

As districts take a more active role in defining and leading efforts to improve classroom instruction, school leaders can lose some of their autonomy in defining what good practice looks like in their own classrooms as well as in allocating staff. Across the Pew Network, districts have taken very different approaches to balancing stronger district direction with school autonomy—from Fayette's initial hands-off stance to Christina's moderate approach to San Diego's and Yonkers' recent efforts at uniform reforms across schools and classrooms. Because these efforts are relatively new, it is too early to make judgments about their ultimate impacts.

We have seen, however, that school flexibility can work for and against district efforts to create a standards-based system. With autonomy over curriculum and instruction, schools—and even teachers within schools—end up doing very different things, with very uneven results.

Moreover, because schools serving poor children often have a disproportionate number of underqualified staff, programs in these schools are likely to be weaker than those in schools with strong leadership and faculty. Yet, unilateral top-down mandates fuel cynicism toward reform and can result in superficial, if any, implementation of the reforms. Superficial implementation of reforms is predictably the case when little professional development and other supports are provided. In the Pew Network sites, the balance of power in curricular, instructional, and staffing decisions between the central office and the school site reflects in part state regulations and tradition. At one extreme, Kentucky legally gives schools final authority over most issues regarding curriculum, instruction, scheduling, and staffing. In the other states, the legal autonomy of schools vis-à-vis the district is less well specified. Most of the sites—Christina, Portland, Pittsburgh, San Diego, and Yonkers—have long traditions of some degree of school autonomy. In District 2, with its history of active district leadership of reform, the tradition has been more mixed. In fact, as Elmore and Burney (1997) describe, District 2 leaders treat different schools differently, depending on their leadership and history of success.

The assumption in Kentucky's state reform has been that accountability plus flexibility would motivate improvement and that teachers could figure out what changes they needed to make. This approach presumed both that teachers would know what to seek in the way of professional development and that it would be available. Neither has turned out to be the case. It also presumed that teachers would be able to develop curriculum aligned to the state assessment. The result has been very uneven educational programs across the district's schools, their quality depending largely on the leadership and faculty at each school. Fayette County leaders, like others throughout Kentucky, have been hesitant to tread on schools' autonomy. The new leadership in Fayette County, as we described above, is developing ways to exercise leadership in a decentralized system, through guidance and support rather than mandates.

At the other extreme, both San Diego and Yonkers are cases where new district leadership has taken an aggressive stance to ensure districtwide compliance with a common curriculum and set of instructional approaches. We discussed the San Diego case in detail earlier in this chapter. It is built on the contention that effective practices are well known and that schools and teachers should put them into place in every classroom. San Diego goes so far as to mandate that a certain amount of time at a certain time of day be focused on balanced literacy approaches in each classroom in the district, including even more time for students at risk of failure. The district also requires principals to be in classrooms for two hours each morning. And the district's hand has been felt strongly in reallocating Title I funds to support this approach, including the decision to stop hiring instructional assistants in order to support staff developers. Not everyone has responded favorably to these new demands—in particular, the teachers' union

has protested teachers' loss of authority, which we discuss in greater detail in the following section. However, unlike other examples of mandated curriculum, the approach to literacy is not often criticized. Criticism is directed instead to the top-down demands that affect staff and schedules at the school site.

In Yonkers, the then new superintendent moved quickly to assert district leadership over instructional practice. He introduced uniform curricula in the elementary schools in both mathematics and literacy, curricula that he had implemented in his previous district. To signal a districtwide commitment to the new curricula, he closed schools for a week and held professional development workshops for all teachers and administrators. The introduction of new curricula represented a major shift for Yonkers, where traditionally individual schools and, in many cases, individual teachers determined what curriculum materials would be used. To support the implementation of the new curricula, principals were expected to spend more time in classrooms, and central office staff began to conduct "walk-throughs" of schools. The superintendent even went so far as to require block scheduling throughout the district, further limiting school autonomy.

The leadership in Christina has taken a more moderate approach to district-led reforms. There, the district has chosen a set of curricula that staff believe best represent the kind of instruction and learning called for in the state standards and the state assessment. The district professional development offerings have been increased and refocused on a set of activities that are aligned with the new curricula. But final decisions about which curriculum to use, scheduling, and instructional approaches remain at the school level—and in some schools at the classroom level. At this point, some schools have adopted the district's approaches and some have not—although the number of schools coming on board is growing. According to the superintendent, "There really is no other way to get there from here—we cannot just mandate everything."

Again, these different approaches to school autonomy are too recent to judge their long-term impacts. It may be, for example, that over the long run, Christina's moderate approach will prove more effective than the rapid district-led efforts in Yonkers. In the case of San Diego, the possibility of long-term success is increased by the considerable professional development and an approach to literacy that resonates with most educators. Yet, whether grassroots support among teachers and principals will develop over time remains an open question.

The Catch 22 for most district leaders, however, is that there is increasing pressure from the school board and the public to demonstrate achievement gains quickly. Doing so by slowly building grassroots support from the beginning may not be an option in many urban settings.

The Role of Unions and the Community in Supporting Reform

Teacher and principal associations, through negotiated contracts, can have a major effect on district reform plans, as can various sectors of the community, either through their presence on the board of education or at the level of neighborhood-based or other community groups. Districts face an uphill battle in attempting to implement major reforms without the support of the organizations that represent educators and without the backing of parents and the business community.

Where districts have built collaborations with unions, their reform efforts are more likely to be supported, although historical tensions may make this difficult to do in some districts. In San Diego and Yonkers, we have observed teachers' unions resist the changes initiated by a new administration. In Yonkers, teachers staged a strike that quickly turned bitter and focused largely on the superintendent's leadership style. The strike was settled after four days, with each side claiming victory. Block scheduling was dropped, but the teachers' workday was lengthened slightly to allow for more professional development time and an extended-day program. In Pittsburgh and San Diego, selection procedures for lead teachers and staff developers, respectively, have run into problems around the role of seniority in such decisions.

The most far-reaching changes in the district role over the four years of the Pew Network have occurred in San Diego. There, the changes were made possible by the increasingly strong role of the business community. Yet the implementation of the reforms has been slowed somewhat by the reticence of the unions to support changes they say they had no hand in devising, and the long-term sustainability of the reforms is in question because of the lack of involvement of key interest groups.

The current majority on the San Diego school board was elected with the financial and political support of the Chamber of Commerce. The business community then supported the hiring of a new superintendent from outside of education, financed the hiring of the Chancellor of Instruction, supported a major bond initiative that provided the schools with over $1 billion, and continues to provide millions of dollars to support the district's professional development initiative.

In contrast, the associations representing teachers and administrators have taken a strong stance against the reforms—or at least against the centralized decision-making process. The teachers' union has staged public protests against the reforms and fought a long and highly publicized campaign against the district's effort to recruit, screen, and train staff developers for the schools outside of the previously bargained hiring procedures that give preference in such assignments to teachers with the most seniority. The principals' association unsuccessfully

challenged the district's reassignment of principals. It has taken the intervention of the leadership in the local universities to craft compromises and keep the reforms—in somewhat altered forms—on track. Whether the reforms will be sustained and the leadership remain in place long enough for the early signs of success to take root and spread remains an ongoing question in light of the opposition of these key groups.

Portland provides an example of influence from another sector of the community on the development of the standards-based reform agenda. There, the continued low performance in a set of schools in the north and northeast sectors of the city led to a threatened boycott of the schools by members of the African-American community. That community called for a more aggressive plan to improve achievement of poor and minority students. The result was the creation of the Community Monitoring Advisory Committee, which was given freedom to conduct monitoring visits to schools to assess their progress. The district also agreed to report quarterly on the progress of poor and minority students at each school and to describe specific actions taken to improve their instruction.

In response to concerns about school quality, the district put in place a procedure for identifying and offering support to low-performing schools. The district then called for schools that continued to perform poorly to be reconstituted. The teachers' union disagreed with the strategy and, on the basis that no clear reconstitution policy existed, urged teachers not to participate in any efforts to reconstitute schools. The union persuaded its membership not to participate on teams charged with supporting low-performing schools. When a large high school was reconstituted, the union criticized the action as sudden and based on unclear and poorly publicized criteria. In response, no Portland teacher applied for a job at the school. As a result, the school had to hire 40 beginning teachers and 30 teachers with experience outside the district.

In Yonkers, the reform agenda continues to unfold in the context of court oversight brought on by a lawsuit alleging unequal educational opportunities for African-American youngsters in the district. Every effort to improve the schools has to be justified in light of its potential impact on minority students in the district. For example, the most recent reform program—an ambitious effort to get every school in the district to adopt new mathematics and literacy curricula—was challenged by the NAACP in court before it could be implemented.

In each of these cases, organized groups outside of the central office have pressured districts to move forward with certain reform efforts while other groups have fought to slow reforms or move them in different directions.

Conclusions

Where we have seen districtwide changes in instruction, we find a strong district vision of effective instructional practice and a comprehensive professional development program and staffing that support that vision. However, to significantly expand and revamp professional development, districts need to make major shifts in priorities, budgets, and staffing. Because these changes have tremendous financial and political ramifications, they are not attractive to many district leaders.

This stronger district role then creates tensions between traditions of school autonomy and centralized control of decision-making. We find that it is districts, not schools, that create districtwide priorities and expectations; and districts make significant choices about the resources available for professional development. In all the sites, we have seen efforts on the part of districts to take more control over the shape and direction of the reform agenda. In fact, we did not find any instances in which schools on a widespread basis were able to make significant improvements in classroom practice in the absence of active support and leadership from the district.

Yet our experience in the Pew Network sites underscores the difficulty of central offices' *making* deep change happen at the classroom level. Teachers and principals retain a degree of autonomy, regardless of the rigidity of district mandates; and teachers and principals, as well as community members, can derail reforms with which they disagree. Without ownership and flexibility to adapt to unique circumstances, principals and teachers may not be motivated to change practice, certainly over the long haul. When district mandates for instruction are put in place quickly, ownership by teachers and principals may be sacrificed. On the other hand, defining basic expectations for acceptable practice and redirecting resources to ensure that all teachers are able to meet those expectations are not likely to happen without strong central direction. These findings raise fundamental questions about a central tenet of the theory of standards-based reform: in return for accountability for results, school staff should be free to achieve those results in the way they deem most appropriate. The rationale behind this piece of the theory is that, given clear learning goals, an aligned assessment system, and a strong accountability system, schools should have latitude to decide how to meet those goals.

What we have found, however, is a much less clear-cut situation. As we described earlier, ever-changing and multiple sets of standards, assessments, and accountability do not provide the clear and steady guidance by which the theory presumes schools can chart their own course. Moreover, regardless of the clarity of standards and their alignment with assessment and accountability, school staff cannot make fundamental changes in their instructional practice in

the absence of significant professional development, which typically must be led from the district. District-led reform, however, by definition reduces the autonomy of school staff. And district-led reforms that do not have the support of teachers and parents are unlikely to survive. Striking the appropriate balance between district and school control is even harder in a political climate that calls for immediate results.

V. CONCLUSIONS

The theory of standards-based reform takes the creation of ambitious standards, aligned assessments, and accountability as the starting place for increasing student achievement. The logic is that these reform components will communicate clear and high expectations for students, guide changes in practice, and motivate educators to improve. The experiences of the Pew Network districts over the last four years suggest that this logic does not hold up well in practice.

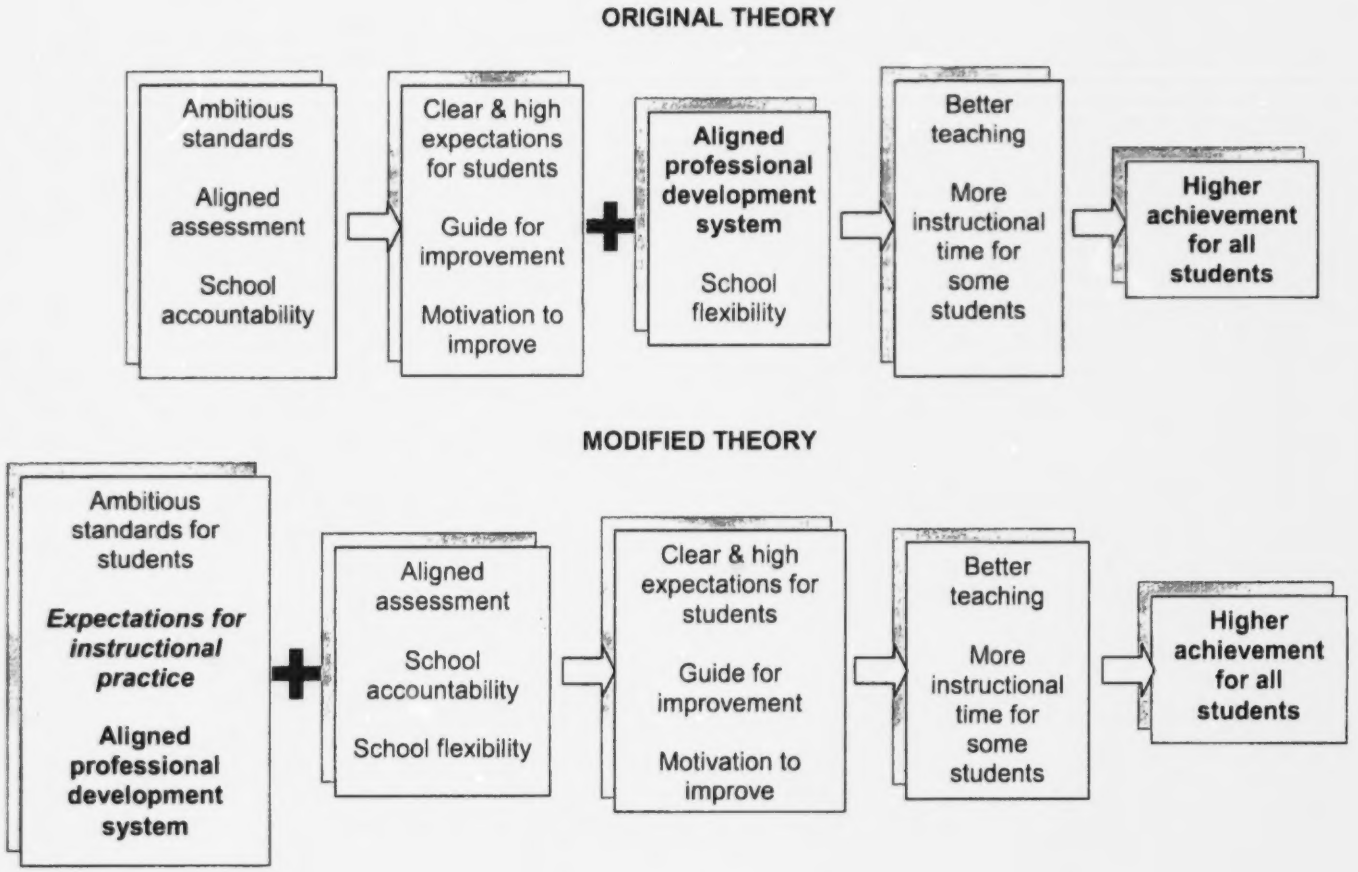
We have seen that districts can have standards, assessments, and accountability in place, yet not improve the quality of curriculum and instruction in classrooms. In contrast, districts that communicate ambitious expectations for instruction, supported by a strong professional development system, are able to make significant changes in classroom practices.

Therefore, we offer a modification of the theory that incorporates what we have learned across the seven Pew Network sites. In the modified theory, shown in Figure 4, ambitious standards for students remain the starting place for improving instruction and hence student achievement. We then add the central idea of high expectations for instructional practice, and we move professional development to the front.

The logic is similar to the original theory. Standards are key to the extent that they focus attention on high expectations for *all* students and communicate the quality of student work that meets the standards. Knowing what is expected of students, however, is not the same as knowing how to help students reach high standards. To help students meet standards, teachers need an explicit picture of what ambitious curriculum and instructional practice look like, and a system of professional development and support that helps them put good ideas into practice. Therefore, we place more emphasis on expectations for instruction and on professional development than on assessment and accountability.

Shifting the order of the boxes in our representation of the theory does not solve the problems districts encounter in translating the ideas of standards-based reform into actions. In fact, such a linear representation of reform provides an overly simplistic view of a set of complex relationships. All the pieces are interrelated, and reform never moves in a clear path from standards to improvement in teaching and learning. In practice, reforms always are initiated in the context of ongoing improvement efforts, changes in leadership, crises of various sorts, budget fluctuations, and changes in state policy. Each of these has a greater influence on what happens next than does the logic of a theory of action. Moreover, districts rarely have a unified reform

Figure 4
The Original and Modified Theories of Standards-Based Reform



94

strategy; more often, districts have multiple efforts that change as leaders, state policies, and funding sources for particular reforms change.

What the modified theory does, however, is draw attention to the importance of what teachers do. Moving expectations for instruction and a supporting system of professional development to the front shifts the focus of attention from assessment and accountability to what it takes to strengthen teaching and learning. If districts take on standards-based reform without a focus on curriculum and instruction, teaching—and, therefore, student learning—is unlikely to change.

Emphasizing the importance of expectations for instruction does not make the task easy for urban districts. They still confront all the issues enumerated in this report: creating an effective instructional guidance and support system with limited resources, building an assessment and accountability system that motivates change in a positive direction, and finding the right balance of direction and autonomy vis-à-vis the schools and the state. And they cannot seriously tackle any of these challenges without public and political support.

Building an Infrastructure for Changes in Classrooms

Creating and communicating clear expectations for instruction and building an effective system of professional development require considerable resources, both human and financial. Building the capacity of schools to design rich learning environments for teachers and students requires building an analogous capacity to support schools in the central office. Teachers are not the only players in districts who need to learn new skills and knowledge about teaching and learning. District leaders and central office staff, as well as principals, need some basis for holding shared understandings about expectations for teachers and students, yet few in the central office have both the instructional knowledge and the skills needed for working with principals and teachers. District leaders who have created the richest learning opportunities for teachers simultaneously created such learning opportunities for central office staff so they could work effectively with principals and teachers.

Building the capacity to lead and support instructional change at all levels of the system is a huge and expensive task. It often requires dismantling existing fiefdoms and fundamentally changing the roles of staff. It is little wonder that districts facing large deficits view the task as impossible. Yet, we observed examples of substantial reallocation of resources toward professional development staffing and activities, made possible by redirecting and combining funding streams from a variety of sources.

Where teacher practice has changed beyond a few teachers or schools, districts have consolidated and targeted funds to teacher support. They also have focused on one or at most two subject areas, and often on a limited grade span. Changes appear more likely on a broad scale when the expectations for instruction make sense to teachers, although such sense making may follow rather than precede learning and trying out new approaches. When students are more engaged and make more progress as a result of new teaching practices, teachers usually embrace the new practices.

Creating Assessment and Accountability Systems That Motivate Improvement

Along with the adoption of standards, assessment and accountability have received the lion's share of the attention in the districts we studied over the last four years. Yet, few have been able to put standards-based assessments in place; most continue to rely in large part on norm-referenced multiple-choice tests. Technical, political, and financial barriers have stymied both state and local efforts to create richer assessments, although some states and districts have made headway. Without such assessments, however, the goal of achieving high standards is undercut.

When sanctions are tied to assessments, teachers are motivated to raise test scores and therefore feel pressed to spend time on test preparation. In districts where multiple-choice tests dominate, test preparation means less time for more ambitious instruction, especially for low-performing students. In districts where assessments reasonably match challenging standards, testing can influence what teachers do in a positive direction. High-stakes assessments also focus attention on the short run, as each year's results present different patterns of strengths and weaknesses.

Accountability systems focus primarily on identifying failing schools and students. Few districts, or their states, have systems that target prevention and provide intensive assistance to schools and students at risk of failure. And fewer define accountability as a two-way street, where levels of the system have mutual responsibilities to each other. Creating systems in which accountability comes from shared responsibility for providing the conditions necessary for ambitious teaching and for improved student results is a direction worth pursuing, but one that is undercut by accountability based on distrust. Districts and states face the challenge of designing accountability systems that do in fact penalize those who have abdicated their responsibility to students without simultaneously undercutting the majority of educators who are motivated to do a good job with the right conditions and assistance.

Balancing Mandates and Discretion

Sitting between schools and the state, districts are constrained in what they can do. As the states have become more active in standards-based reform, the Pew Network districts have found that they are not necessarily marching in the same direction as the state. The vision that states would provide clear and steady guidance through standards and assessments within which districts would operate has not come to pass. And given the historical tension between large cities and their states, it is not surprising that dual systems of standards and assessments have emerged in many cases.

Within districts, similar tensions are created when districts drive reform with a heavy hand, particularly when schools have been accustomed to considerable discretion over certain resources. District leaders are challenged to lead in ways that provide strong direction to schools without leaving school faculties feeling disenfranchised. Some degree of flexibility and professional discretion is essential for principals and teachers; the question is, how much and within what constraints. Pressures on district leaders for quick results—from their states and their local boards—make the efficiency of mandates appealing. But mandates without teacher buy-in may not grow deep enough roots to survive long enough to make a difference.

Continuity in direction also challenges both state and district leaders. Whether through state action or change in boards or superintendents, standards-based reform cannot move forward when dramatic shifts in direction take place every few years. Nor can it move forward without continually revisiting and revising standards and assessments. The key is finding the right balance between continuity in direction and continuous improvement.

Promising Trends

Creating an infrastructure to support teacher learning, rich and constructive assessment and accountability, and the right balance of authority and discretion present significant challenges to district leaders. The fact that few of the Pew Network districts made substantial progress is testimony to the difficulty of the task. Still, we saw signs that educators were taking reform seriously everywhere and, across all the districts, observed several trends over the four years of the evaluation that bode well for continued progress. These trends include:

- Substantially more attention to professional development for teachers, including placement of staff developers in schools.
- A shift away from focusing on a few schools toward including all schools in reform efforts.

- Expansion of assessment systems to incorporate some form of testing that requires demonstration of work beyond checking one of several choices.
- An increase in attention to and use of classroom-based diagnostic assessment with the primary purpose of informing instruction.
- Greater emphasis on supporting principals to become instructional leaders.
- Increased attention to data in school planning.
- Examples of richer notions of accountability that rely on multiple measures, professional judgment, and shared responsibility for student learning.
- More attention to the importance of a district strategy for change that focuses attention on one or two subject areas.
- More opportunities for students who are failing or who are at risk of failure to have extra instruction that is challenging, not remedial.

These trends are fragile, but they have the potential to continue and even accelerate if investments are forthcoming in creating the human capital necessary for reform. Great as the challenges are in putting the basic building blocks for standards-based reform in place, they go beyond those identified above. For standards-based reform rests on a premise—that all children can achieve at high levels—radically different from the one that has traditionally guided schools in the United States. This view rejects the long-held belief that student achievement will always be normally distributed, with some high achievers, most students in the middle, and some who will not succeed. Instead, standards-based reform calls for setting ambitious benchmarks for learning for all students and providing the instruction necessary to get them there. In so doing, it challenges deeply held beliefs about who can and who cannot succeed.

The call to increase the learning of all students runs counter to many current trends, including the tendency for less-qualified and less-experienced teachers to be in the lowest-performing schools. Taking the call seriously requires extra efforts to help students who start their formal schooling with weak backgrounds or who fall behind at various points along the way—facts that are true disproportionately for children of poverty and of color. Moreover, a commitment to improving instruction for all students inevitably means concentrating more resources in the schools and classrooms with students who are farthest from reaching the standards. Here is where the rhetoric of standards-based reform runs most directly into entrenched interests in maintaining the status quo.

The bottom line is that taking standards-based reform seriously has profound implications for the priorities, organization, and resource allocation of school districts. Difficult as it is to create ambitious standards and rich assessments, it is far more difficult for districts to equip

central office and school staff with the knowledge and skills needed for their new roles. Yet without these changes, the promise of standards-based reform to increase learning for all students will not be realized. Across the Pew Network districts, the greatest strides occur where the adults also have opportunities to learn.

REFERENCES

Briars, D. J., & Resnick, L. B. (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement.* CSE Technical Report 528. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.

Elmore, R. F., & Burney, D. (1997). *School variation and systemic instructional improvement in Community School District #2, New York City.* Pittsburgh, PA: University of Pittsburgh, HPLC Project, Learning Research and Development Center.

Elmore, R. F., & Rothman, R. (Eds.) (1999). *Testing, teaching, and learning: A guide for states and school districts.* Washington, DC: National Academy Press.

National Center on Education and the Economy. (1990). *America's choice: High skills or low wages. The Report of The Commission on the Skills of the American Workforce.* Rochester, NY: Author.

Neufeld, B. (2000). *Implementing standards-based reform in San Diego City Schools: Update report.* Cambridge, MA: Education Matters, Inc.

Shields, P. M., David, J. L., Humphrey, D. C., & Young, V. M. (1999). *Pew Network for Standards-Based Reform: Year three evaluation report.* Menlo Park, CA: SRI International.

Smith, M. S., & O'Day, J. (1990). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). Bristol, PA: Falmer Press.

Stein, M. K., Harwell, M., & D'Amico, L. (1999). *Toward closing the gap in literacy instruction.* Pittsburgh, PA: University of Pittsburgh, HPLC Project, Learning Research and Development Center.

APPENDIX

CHARACTERISTICS OF PEW NETWORK DISTRICTS

Table A-1
Size of Pew Network Districts

District	Number of Students	Number of Schools		
		Elementary	Middle	High
Christina	20,039	19	3	3
District 2	22,640	28	15	0
Fayette	32,445	35	11	5
Pittsburgh	38,846	60	21	12
Portland	54,427	63	17	10
San Diego	142,326	124	23	17*
Yonkers	25,935	30	4	5

Note: All data from district Web sites 1/20/01 (except Christina from 1995 Pew data).
* Plus 13 multi-level schools.

Table A-2
Demographics of Pew Network Districts

District	Ethnicity (Percent)					Percent LEP	Percent Free/ Reduced-Price Lunch
	Caucasian	African-American	Latino	Asian	Other Minority		
Christina	62	30	4	*	4	n/a***	32
District 2	32	14	20	32	0	13	48
Fayette	75	23	*	*	3	n/a	32
Pittsburgh	41	57	*	*	2	n/a	52
Portland	62	17	9	10	2	6	n/a
San Diego	28	17	37	3	15**	50	75
Yonkers	28	30	39	*	3	n/a	66

Note: All data from district Web sites 1/20/01 (except Christina and Pittsburgh free/reduced-price lunch from 1995 Pew data).
*Reported as part of "Other Minority."
**Includes 8% Filipino and 6% Indochinese.
***Not Available

Endnotes

[1] For purposes of this final report, we needed to construct a statement of the theory of standards-based reform that was consistent with standards-based policies and language and would serve the practical purpose of structuring our main findings. Therefore, our articulation of the theory aims to make explicit the "theory of action" implied by much of the standards-based education reform legislation in the1990s. Examples include Goals 2000, the NSF Systemic Initiatives, and the 1994 reauthorization of Title I (of the Elementary and Secondary Education Act of 1965) at the federal level. At the state level, examples are Kentucky's Education Reform Act of 1990 and Oregon's Educational Act of the 21st Century of 1992. We also drew on conversations with staff at the Pew Charitable Trusts and on a variety of reports, particularly the National Research Council's Title I Guide (*Testing, Teaching, and Learning*), which describes the theory of action behind the Title I legislation (Elmore & Rothman, 1999).

[2] Most test publishers of nationally normed tests now offer to report their results in terms of "standards." Typically, this means merely that they have defined cut scores for categories such as "proficient" or "advanced," which may or may not reflect the performance expectations of a given state or district. In addition, many states and districts treat the 50th percentile as the defining standard for accomplishment and often report norm-referenced scores in terms of percent who "meet" that standard, without regard to the relationship of percentiles to their own particular standards and expectations for students.

[3] District 2 has been documented in far more detail than we provide by Elmore and Burney, as well as by other researchers involved in the High Performance Learning Communities project. See http://www.lrdc.pitt.edu/hplc for these reports.

[4] We visited only a few schools, but our conversations with other researchers who have visited many schools suggest a similar pattern: even when teachers and principals are antagonistic to the reforms, they understand the goals of the reforms. In fact, most educators are positive about the content of the Literacy Framework and the associated professional development; "it's the how, not the what" that has resulted in resistance. See, for example, Neufeld (2000).

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

UD 035 840

I. DOCUMENT IDENTIFICATION:

Title: When Theory Hits Reality: Standards-Based Reform in Urban Districts

Author(s):

Corporate Source: SRT International for The New Charitable Trusts.

Publication Date: August 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to each document.

If permission is granted to reproduce and disseminate the identified documents, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

SAMPLE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

SAMPLE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

SAMPLE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate these documents as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature:	Printed Name/Position/Title: Sue Unger - Director, Education
Organization/Address: The New Charitable Trusts	Telephone: 215 574 4755 FAX: 215 574 4537
	E-Mail Address: sunger@pewtrusts.com Date: 8/13/03